Coordinate Descent

Nipun Batra

IIT Gandhinagar

July 30, 2025

where:

$$\hat{y}_i^{(-j)} = \theta_0 x_i^0 + \ldots + \theta_d x_i^d$$

is $\hat{y}_i$ without θ_j

Coordinate Descent for Unregularized regression

$$Set \frac{\partial \, \mathrm{RSS}\left(\theta_j\right)}{\partial \theta_j} = 0$$

$$\theta_j = \sum_{i=1}^{N} \frac{\left(y_i - \left(\theta_0 x_i^0 + \ldots + \ldots + \theta_d x_i^d\right)\right)\left(x_i^j\right)}{\left(x_i^j\right)^2} = \frac{\rho_j}{z_j}$$

$$\rho_j = \sum_{i=1}^{N} x_i^j \left(y_i - \hat{y}_i^{(-j)})\right)$$

$$z_j = \sum_{i=1}^{N} \left(x_i^j\right)^2$$

z_j is the squared of ℓ_2 norm of the j^{th} feature

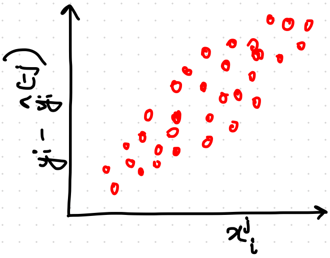
UNDERSTANDING ρ_j IN COORDINATE DESCENT

$$\rho_j = \sum_{i=1}^{N} x_i^j \left(y_i - \hat{y}_i^{(-j)} \right)$$

UNDERSTANDING ρ_j IN COORDINATE DESCENT

$$\rho_j = \sum_{i=1}^{N} x_i^j \left(y_i - \hat{y}_i^{(-j)} \right)$$

CASE 1

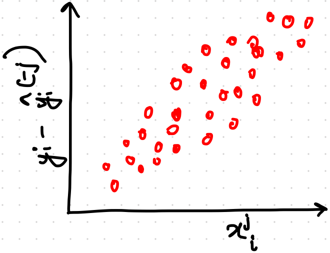


x_i^j STRONG +VE CORR. WITH $y_i - \hat{y}_i^{(-j)}$

UNDERSTANDING ρ_j IN COORDINATE DESCENT

$$\rho_j = \sum_{i=1}^{N} x_i^j \left(y_i - \hat{y}_i^{(-j)}\right)$$

CASE 1



x_i^j STRONG +VE CORR. WITH $y_i - \hat{y}_i^{(-j)}$

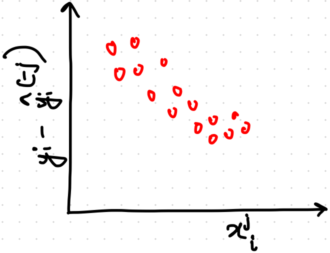
$\Downarrow$

j^{th} FEATURE IS IMPT. AND ITS COEFFICIENT +VE

UNDERSTANDING ρ_j IN COORDINATE DESCENT

$$\rho_j = \sum_{i=1}^{N} x_i^j \left(y_i - \hat{y}_i^{(-j)}\right)$$

CASE II



x_i^j STRONG $-$VE CORR. WITH $y_i - \hat{y}_i^{(-j)}$

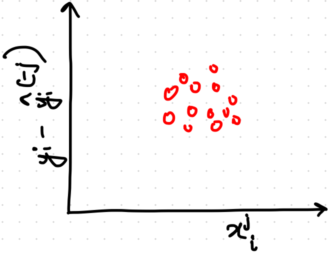
$$\Downarrow$$

j^{th} FEATURE IS IMPT. AND ITS COEF. $-$VE

UNDERSTANDING ρ_j IN COORDINATE DESCENT

$$\rho_j = \sum_{i=1}^{N} x_i^j \left(y_i - \hat{y}_i^{(-j)}\right)$$

CASE II



x_i^j WEAK CORR. WITH $y_i - \hat{y}_i^{(-j)}$

$\Downarrow$

j^{th} FEATURE IS NOT IMPT. AND ITS COEF. $\rightarrow 0$

Coordinate Descent for Lasso Regression

$$\text{Minimize} \underbrace{\sum_{i=1}^{N} \epsilon^2 + \delta^2 \left\{ |\theta_0| + |\theta_1| + \ldots \left|\theta_j\right| + \ldots |\theta_d| \right\}}_{\textit{LASSO OBJECTIVE}}$$

$$\frac{\partial}{\partial \theta_j}(\text{LASSO OBJECTIVE}) = -2\rho_j + 2\theta_j z_j + \delta^2 \frac{\partial}{\partial \theta_j} \left|\theta_j\right|$$

$$\frac{\partial}{\partial \theta_j} \left|\theta_j\right| = \begin{cases} 1 & \theta_j > 0 \\ [-1, 1] & \theta_j = 0 \\ -1 & \theta_j < 0 \end{cases}$$

Coordinate Descent for Lasso Regression

- **Case 1:** $\theta_j > 0$

$$
\begin{aligned}
2\rho_j + 2\theta_j z_j + \delta^2 &= 0 \\
\theta_j &= \frac{\rho_j - \frac{\delta^2}{2}}{z_j} \\
\rho_j > \frac{\delta^2}{2} \Rightarrow \theta_j &= \frac{\rho_j - \frac{\delta^2}{2}}{z_j}
\end{aligned}
$$

Coordinate Descent for Lasso Regression

- **Case 1:** $\theta_j > 0$

$$2\rho_j + 2\theta_j z_j + \delta^2 = 0$$

$$\theta_j = \frac{\rho_j - \frac{\delta^2}{2}}{z_j}$$

$$\rho_j > \frac{\delta^2}{2} \Rightarrow \theta_j = \frac{\rho_j - \frac{\delta^2}{2}}{z_j}$$

Coordinate Descent for Lasso Regression

- **Case 1:** $\theta_j > 0$

$$2\rho_j + 2\theta_j z_j + \delta^2 = 0$$
$$\theta_j = \frac{\rho_j - \frac{\delta^2}{2}}{z_j}$$
$$\rho_j > \frac{\delta^2}{2} \Rightarrow \theta_j = \frac{\rho_j - \frac{\delta^2}{2}}{z_j}$$

- **Case 2:** $\theta_j < 0$

$$\rho_j < \frac{\delta^2}{2} \Rightarrow \theta_j = \frac{\rho_j + \delta^2/2}{z_j} \tag{1}$$

Coordinate Descent for Lasso Regression

- **Case 3:** $\theta_j = 0$

$$\frac{\partial}{\partial \theta_j}(\text{LASSO OBJECTIVE}) = -2\rho_j + 2\theta_j z_j + \delta^2 \underbrace{\frac{\partial}{\partial \theta_j}|\theta_j|}_{\text{[-1,1]}}$$

$$\in \underbrace{[-2\rho_j - \delta^2, -2\rho_j + \delta^2]}_{\{0\}\text{ lies in this range}}$$

$$-2\rho_j - \delta^2 \leq 0 \text{ and } -2\rho_j - \delta^2 \leq 0$$

$$-\frac{\delta^2}{2} \leq \rho_j \leq \frac{\delta^2}{2} \Rightarrow \theta_j = 0$$

Summary of Lasso Regression

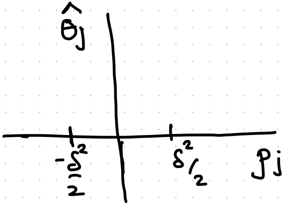
$$\theta_j = \begin{bmatrix} \frac{\rho_j + \frac{\delta^2}{2}}{z_j} & \textit{if} & \rho_j < -\frac{\delta^2}{2} \\ 0 & \textit{if} & -\frac{\delta^2}{2} \leq \rho_j \leq \frac{\delta^2}{2} \\ \frac{\rho_j - \frac{\delta^2}{2}}{z_j} & \textit{if} & \rho_j > \frac{\delta^2}{2} \end{bmatrix} \tag{2}$$

LASSO (SOFT) THRESHOLDING

$$\theta_j = \begin{cases} \dfrac{\rho_j + \delta^2/2}{z_j} & \text{if} \quad \rho_j < -\delta^2/2 \\ 0 & \text{if} \quad -\dfrac{\delta^2}{2} \le \rho_j \le \dfrac{\delta^2}{2} \\ \dfrac{\rho_j - \delta^2/2}{z_j} & \text{if} \quad \rho_j > \delta^2/2 \end{cases}$$

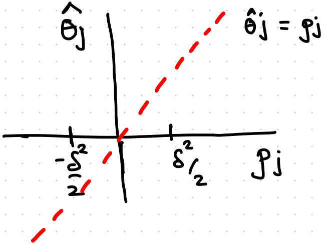
LASSO (SOFT) THRESHOLDING

$$\hat{\theta}_j = \begin{cases} \dfrac{\rho_j + \delta^2/2}{z_j} & \text{if} \quad \rho_j < -\delta^2/2 \\ 0 & \text{if} \quad -\dfrac{\delta^2}{2} \le \rho_j \le \dfrac{\delta^2}{2} \\ \dfrac{\rho_j - \delta^2/2}{z_j} & \text{if} \quad \rho_j > \delta^2/2 \end{cases}$$



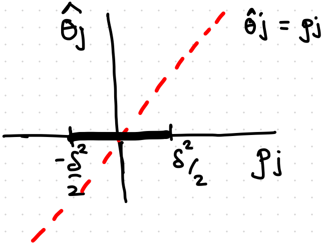
LASSO (SOFT) THRESHOLDING

$$\hat{\theta}_j = \begin{cases} \dfrac{\rho_j + \delta^2/2}{z_j} & \text{if} \quad \rho_j < -\delta^2/2 \\ 0 & \text{if} \quad -\dfrac{\delta^2}{2} \le \rho_j \le \dfrac{\delta^2}{2} \\ \dfrac{\rho_j - \delta^2/2}{z_j} & \text{if} \quad \rho_j > \delta^2/2 \end{cases}$$



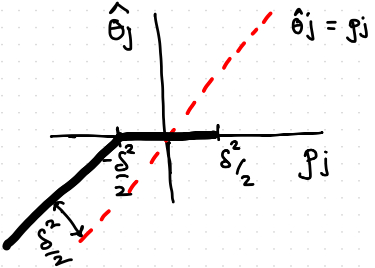
LASSO (SOFT) THRESHOLDING

$$\hat{\theta}_j = \begin{cases} \dfrac{\rho_j + \delta^2/2}{z_j} & \text{if} \quad \rho_j < -\delta^2/2 \\ 0 & \text{if} \quad -\dfrac{\delta^2}{2} \le \rho_j \le \dfrac{\delta^2}{2} \\ \dfrac{\rho_j - \delta^2/2}{z_j} & \text{if} \quad \rho_j > \delta^2/2 \end{cases}$$



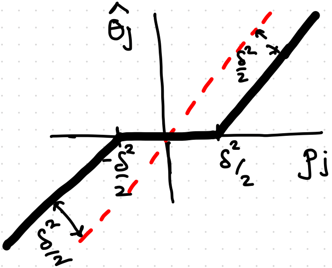
LASSO (SOFT) THRESHOLDING

$$\hat{\theta}_j = \begin{cases} \dfrac{\rho_j + \delta^2/2}{z_j} & \text{if} \quad \rho_j < -\delta^2/2 \\ 0 & \text{if} \quad -\dfrac{\delta^2}{2} \le \rho_j \le \dfrac{\delta^2}{2} \\ \dfrac{\rho_j - \delta^2/2}{z_j} & \text{if} \quad \rho_j > \delta^2/2 \end{cases}$$



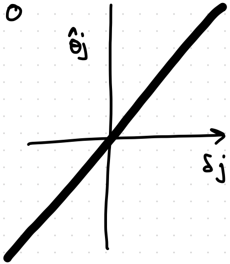
LASSO (SOFT) THRESHOLDING

$$\hat{\theta}_j = \begin{cases} \dfrac{\rho_j + \delta^2/2}{z_j} & \text{if} \quad \rho_j < -\delta^2/2 \\ 0 & \text{if} \quad -\dfrac{\delta^2}{2} \le \rho_j \le \dfrac{\delta^2}{2} \\ \dfrac{\rho_j - \delta^2/2}{z_j} & \text{if} \quad \rho_j > \delta^2/2 \end{cases}$$



LASSO (SOFT) THRESHOLDING

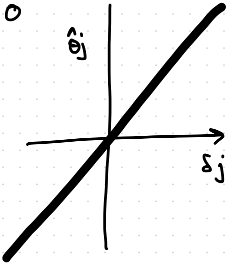
$\delta^2 = 0$

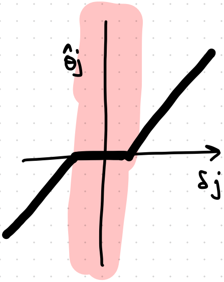


NO REGULARISATION

LASSO (SOFT) THRESHOLDING

$\delta^2 = 0$





NO REGULARISATION

REGULARISATION

$\Downarrow$

SPARSITY